

MSC: 68W25, 68Q25

DOI: 10.21538/0134-4889-2017-23-3-159-170

APPROXIMATION SCHEME FOR THE PROBLEM OF WEIGHTED 2-PARTITIONING WITH A FIXED CENTER OF ONE CLUSTER

A. V. Kel'manov, A. V. Motkova, V. V. Shenmaier

We consider the intractable problem of partitioning a finite set of points in Euclidean space into two clusters with minimum sum over the clusters of weighted sums of squared distances between the elements of the clusters and their centers. The center of one cluster is unknown and is defined as the mean value of its elements (i.e., it is the centroid of the cluster). The center of the other cluster is fixed at the origin. The weight factors for the intracluster sums are given as input. We present an approximation algorithm for this problem, which is based on the adaptive grid approach to finding the center of the optimal cluster. We show that the algorithm implements a fully polynomial-time approximation scheme (FPTAS) in the case of fixed space dimension. If the dimension is not fixed but is bounded by a slowly growing function of the number of input points, the algorithm realizes a polynomial-time approximation scheme (PTAS).

Keywords: Euclidean space, partitioning, NP-hardness, FPTAS, PTAS.

REFERENCES

1. Kel'manov A. V., Romanchenko S. M. An FPTAS for a vector subset search problem. *J. Appl. Ind. Math.*, 2014, vol. 8, no. 3, pp. 329–336. doi: 10.1134/S1990478914030041.
2. Kel'manov A. V., Khandeev V. I. Fully polynomial-time approximation scheme for a special case of a quadratic euclidean 2-clustering problem. *Comput. Math. Math. Phys.*, 2016, vol. 56, no. 2, pp. 334–341. doi: 10.1134/S0965542516020111.
3. Kel'manov A. V., Motkova A. V. A fully polynomial-time approximation scheme for a special case of a balanced 2-clustering problem. Proc. Internat. Conf. on Discrete Optimization and Operations Research (DOOR 2016), 2016, *Ser. Lecture Notes Comp. Sci.*, vol. 9869, pp. 182–192. doi: 10.1007/978-3-319-44914-2–15.
4. Aggarwal A., Imai H., Katoh N., Suri S. Finding k points with minimum diameter and related problems. *J. Algorithms*, 1991, vol. 12, no. 1, pp. 38–56. doi: 10.1016/0196-6774(91)90022-Q.
5. Kel'manov A. V., Pyatkin A. V. NP-completeness of some problems of choosing a vector subset. *J. Appl. Ind. Math.*, 2011, vol. 5, no. 3, pp. 352–357. doi: 10.1134/S1990478911030069.
6. Shenmaier V. V. Solving some vector subset problems by Voronoi diagrams. *J. Appl. Industr. Math.*, 2016, vol. 10, no. 4, pp. 560–566. doi: 10.1134/S199047891604013X.
7. Kel'manov A. V., Romanchenko S. M. Pseudopolynomial algorithms for certain computationally hard vector subset and cluster analysis problems. *Automation and Remote Control*, 2012, vol. 73, no. 2, pp. 349–354. doi: 10.1134/S0005117912020129.
8. Kel'manov A. V., Romanchenko S. M. An approximation algorithm for solving a problem of search for a vector subset. *J. Appl. Ind. Math.*, 2012, vol. 6, no. 1, pp. 90–96. doi: 10.1134/S1990478912010097.
9. Shenmaier V. V. An approximation scheme for a problem of search for a vector subset. *J. Appl. Ind. Math.*, 2012, vol. 6, no. 3, pp. 381–386. doi: 10.1134/S0081543816090066.
10. Gimadi E. Kh., Kel'manov A. V., Kelmanova M. A., Khamidullin S. A. A posteriori detection of a quasiperiodic fragment with a given number of repetitions in a numerical sequence. *Sib. Zh. Ind. Mat.*, 2006, vol. 9, no. 1, pp. 55–74 (in Russian).
11. Gimadi E. Kh., Kel'manov A. V., Kel'manova M. A., Khamidullin S. A. A posteriori detecting a quasiperiodic fragment in a numerical sequence. *Pattern Recognition and Image Analysis*, 2008, vol. 18, no. 1, pp. 30–42. doi:10.1134/S1054661808010057.

12. Kel'manov A. V., Pyatkin A. V. Complexity of certain problems of searching for subsets of vectors and cluster analysis. *Comput. Math. Math. Phys.*, 2009, vol. 49, no. 11, pp. 1966–1971. doi: 10.1134/S0965542509110128.
13. Baburin A.E., Gimadi E.Kh., Glebov, N.I., Pyatkin, A.V. The problem of finding a subset of vectors with the maximum total weight. *J. Appl. Industr. Math.*, 2008, vol. 2, no. 1, pp. 32–38. doi: 10.1007/s11754-008-1004-3.
14. Gimadi E. Kh., Pyatkin A. V., Rykov I. A. On polynomial solvability of some problems of a vector subset choice in a Euclidean space of fixed dimension. *J. Appl. Indust. Math.*, 2010, vol. 4, no. 1, pp. 48–53. doi: 10.1134/S1990478910010084.
15. Gimadi E.K., Glazkov Y.V., Rykov I.A. On two problems of choosing some subset of vectors with integer coordinates that has maximum norm of the sum of elements in Euclidean space. *J. Appl. Ind. Math.*, 2009, vol. 3, no. 3, pp. 343–352. doi: 10.1134/S1990478909030041.
16. Kel'manov A. V., Khandeev V. I. An exact pseudopolynomial algorithm for a problem of the two-cluster partitioning of a set of vectors. *J. Appl. Ind. Math.*, 2015, vol. 9, no. 4, pp. 497–502. doi: 10.1134/S1990478915040067.
17. Dolgushev A. V., Kel'manov A. V. An approximation algorithm for solving a problem of cluster analysis. *J. Appl. Indust. Math.*, 2011, vol. 5, no. 4, pp. 551–558. doi: 10.1134/S1990478911040107.
18. A.V. Dolgushev, A.V. Kel'manov, V.V. Shenmaier. Polynomial-time approximation scheme for a problem of partitioning a finite set into two clusters. *Proc. Steklov Inst. Math.*, 2016, vol. 295, suppl. 1, pp. 47–56. doi: 10.1134/S0081543816090066.
19. Kel'manov A. V., Khandeev V. I. A Randomized algorithm for two-cluster partition of a set of vectors. *Comput. Math. Math. Phys.*, 2015, vol. 55, no. 2, pp. 330–339. doi: 10.1134/S096554251502013X.
20. Kel'manov A. V., Motkova A. V. *J. Appl. Ind. Math.*, 2016, vol. 10, no. 3, pp. 349–355. doi: 10.1134/S1990478916030054.
21. Kel'manov A. V., Pyatkin A. V. NP-hardness of some quadratic euclidean 2-clustering problems. *Dokl. Math.*, 2015, vol. 92, no. 2, pp. 634–637. doi: 10.1134/S1064562415050233.
22. Kel'manov A. V., Pyatkin A. V. On the complexity of some quadratic euclidean 2-clustering problems. *Comput. Math. Math. Phys.*, 2016, vol. 56, no. 3, pp. 491–497. doi: 10.1134/S096554251603009X.
23. Aggarwal C. C. *Data Mining: The Textbook*. Cham, Springer International Publishing, 2015, 734 p. ISBN: 978-3319141411.
24. Bishop C. M. *Pattern Recognition and Machine Learning*. New York: Springer Science+Business Media, LLC, 2006. 738 p. ISBN: 978-0-387-31073-2.
25. Hastie T., Tibshirani R., Friedman J. *The Elements of statistical learning: Data mining, inference, and prediction*. New York: Springer-Verlag. 2009. 763 p. doi: 10.1007/978-0-387-84858-7.
26. James G., Witten D., Hastie T., Tibshirani R. *An introduction to statistical learning*. New York: Springer Science+Business Media, LLC, 2013, 426 p. doi: 10.1007/978-1-4614-7138-7.
27. Jain A. K. Data clustering: 50 years beyond k -means. *Pattern Recognition Lett.*, 2010, vol. 31, no. 8, pp. 651–666. doi: 10.1016/j.patrec.2009.09.011.
28. Wirth N. *Algorithms + data structures = programs*. New Jersey: Prentice Hall, 1976, 366 p. ISBN: 0130224189.
29. Ball K. An elementary introduction to modern convex geometry. Flavors of geometry. *MSRI Publications*, S. Levy editor, 1997, vol. 31, pp. 1–58. ISBN: 0-521-62048-1.

The paper was received by the Editorial Office on May 24, 2017.

Aleksander Vasil'evich Kel'manov, Dr. Phys.-Math. Sci., Sobolev Institute of Mathematics; Novosibirsk State University, Novosibirsk, 630990 Russia, e-mail: kelm@math.nsc.ru.

Anna Vladimirovna Motkova, undergraduate student, Novosibirsk State University, Novosibirsk, 630990 Russia, e-mail: anitamo@mail.ru.

Vladimir Vladimirovich Shenmaier, Cand. Sci. (Phys.-Math.), Sobolev Institute of Mathematics, Novosibirsk, 630990 Russia, e-mail: shenmaier@mail.ru.

Cite this article as:

A. V. Kel'manov, A. V. Motkova, V. V. Shenmaier, Approximation scheme for the problem of weighted 2-partitioning with a fixed center of one cluster, *Trudy Inst. Mat. Mekh. UrO RAN*, 2017, vol. 23, no. 3, pp. 159–170.